



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Otantamenetelmät (78143) Syksy 2010

TEEMA 2

Risto Lehtonen
risto.lehtonen@helsinki.fi



Teema 2

LISÄTIEDON KÄYTTÖ ESTIMOINTIASETELMASSA: MALLIAVUSTEINEN ESTIMOINTI

Risto Lehtonen

2



Lisätiedon käyttö estimointiasetelmassa

- Malliavusteiset strategiat
Model-assisted strategies
 - Regressioestimointi
 - Suhde-estimointi
 - Jälkiosittaminen
 - Kalibrointimenetelmät
- Otoksen ulkopuolisen lisäinfon tuonti
estimointiasetelmaan tilastollisen mallin
avulla
- SAS Procedure SURVEYREG
- SAS makro CLAN

Risto Lehtonen

3



Estimointistrategioita perusjoukon kokonaismäärälle

<i>Strategia</i>	<i>Lisäinformaatio</i>	<i>Avustava malli</i>
Asetelmaperusteisia strategioita		
SRSWOR	Ei ole	Ei ole
SRSWR	Ei ole	Ei ole
STR	Ositusmuuttuja	Ei ole
PPSWOR	Kokomuuttuja	Ei ole
Malliavusteisia strategioita Otanta-asetelma SRS Avustava malli: Lineaarinen kiinteiden tekijöiden malli		
Regressioestimointi SRS*reg	Jatkuva	Regressiomalli
Suhde-estimointi SRS*rat	Jatkuva	Regressiomalli (ei vakiotermiä)
Jälkiositus SRS*pos	Diskreetti	ANOVA

Risto Lehtonen

4



Päämenetelmät

- Lisätiedon käyttö estimointiasetelmassa
- Malliavusteinen estimointi
Model assisted estimation
 - Yleistetyt regressioestimaattorit
Generalized regression (GREG) estimators
- Kalibrointimenetelmät
Calibration
 - Mallista vapaa kalibrointi
Model-free calibration
 - Malliavusteinen kalibrointi
Model calibration

Risto Lehtonen

5



Malliavusteinen estimointi

- Alan klassikko:
Särndal C.-E, Swensson B. and Wretman J. (1992). *Model Assisted Survey Sampling*. Springer.
- Logistinen regressioestimaattori LGREG:
Lehtonen, R. and A. Veijanen (1998). Logistic generalized regression estimators. *Survey Methodology* **24**, 51-55.
- Sovellus osajoukko- ja pienalue-estimointiin:
Lehtonen R. and Veijanen A. (2009). Design-based Methods of Estimation for Domains and Small Areas. In: *Handbook of Statistics 29B; Sample Surveys: Inference and Analysis*. Eds. D. Pfeiffermann and C.R. Rao. North Holland. pp. 219-249.

Risto Lehtonen

6

GREG-estimaattori-perhe

- Mallin valinta GREG-estimaattorille
 - Lineaarinen kiinteiden tekijöiden malli
 - *Linear fixed-effects model*
 - Jatkuva tulosmuuttuja
 - Logistinen kiinteiden tekijöiden malli
 - Binäärinen/moniluokkainen tulosmuuttuja
 - Yleistetty lineaarinen malli
 - *Generalized linear model*
 - Yleistetty lineaarinen sekamalli
 - *Generalized linear mixed model GLMM*
 - Pienalue-estimointi (*small area estimation, SAE*)
 - [Handbook of Statistics](#)

Risto Lehtonen

7

GREG-estimaattori, tekn. tiivistelmä Sovellusalan pienalue-estimointi

$U = \{1, 2, \dots, k, \dots, N\}$ Perusjoukko (äärellinen)

$U_1, \dots, U_d, \dots, U_D$ Kiinnostuksen kohteena olevat pj:n osajoukot

Otanta-asetelma: PPS-WOR, otoskoko n

$s \subset U$ Otos

$s_d = s \cap U_d$ Osajoukkoon d kuuluva otos

$\pi_k = n \frac{x_{1k}}{\sum_{k \in U} x_{1k}}$ Sisältymistn alkion $k \in U$ PPS-otannassa

$a_k = 1/\pi_k$ Asetelmapaino alkion $k \in s$

Tulosmuuttujan y havaitut arvot y_k alkion $k \in s$

Risto Lehtonen

8

· Yleistetty lineaarinen kiinteiden vaikutusten malli (1)

Yhteinen malli kaikille osajoukoille d

$$E_m(y_k) = f(\mathbf{x}'_k \boldsymbol{\beta})$$

missä

$f(\cdot)$ mallin funktionaalinen muoto

$$\mathbf{x}_k = (1, x_{1k}, \dots, x_{pk})', \quad k \in U$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$$

β_j kiinteät termit (yhteisiä kaikille d)

$$j = 0, \dots, p$$

Mallilla lasketut sovitteet $\hat{y}_k = f(\mathbf{x}'_k \hat{\boldsymbol{\beta}})$, $k \in U$

· Yleistetty lineaarinen kiinteiden vaikutusten malli (2)

Osajoukkokohtaiset vakiotermit

$$E_m(y_k) = f(\mathbf{x}'_k \boldsymbol{\beta})$$

missä

$f(\cdot)$ mallin funktionaalinen muoto

$$\mathbf{x}_k = (I_{1k}, \dots, I_{Dk}, x_{1k}, \dots, x_{pk})', \quad k \in U$$

$$I_{dk} = 1 \text{ if } k \in U_d, \quad I_{dk} = 0 \text{ muulloin, } d = 1, \dots, D$$

$$\boldsymbol{\beta} = (\beta_{01}, \dots, \beta_{0D}, \beta_1, \dots, \beta_p)'$$

β_{0d} osajoukkokohtaiset vakiotermit, $d = 1, \dots, D$

β_j yhteiset kiinteät termit, $j = 1, \dots, p$

Mallilla lasketut sovitteet $\hat{y}_k = f(\mathbf{x}'_k \hat{\boldsymbol{\beta}})$, $k \in U$



Yleistetty lineaarinen sekamalli

Osajoukkokohtaiset satunnaistermit

(random intercepts)

$$E_m(y_k | \mathbf{u}_d) = f(\mathbf{x}'_k(\boldsymbol{\beta} + \mathbf{u}_d)), \quad d = 1, \dots, D$$

missä

$f(\cdot)$ mallin funktionaalinen muoto

$$\mathbf{x}_k = (1, x_{1k}, \dots, x_{pk})'$$

$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ kiinteät termit

$\mathbf{u}_d = (u_{0d}, \dots, u_{pd})'$ satunnaistermit

Mallilla lasketut sovitteet $\hat{y}_k = f(\mathbf{x}'_k(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d)), \quad k \in U$



Erikoistapauksia

(1) Logistinen kiinteiden tekijöiden malli

$$E_m(y_k) = \frac{\exp(\mathbf{x}'_k \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_k \boldsymbol{\beta})}$$

Sovitteet $\hat{y}_k = \frac{\exp(\mathbf{x}'_k \hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{x}'_k \hat{\boldsymbol{\beta}})}, \quad k \in U$

(2) Lineaarinen sekamalli, satunnaiset vakiotermit

$$E_m(y_k | \mathbf{u}_d) = \mathbf{x}'_k \boldsymbol{\beta} + u_{0d}, \quad d = 1, \dots, D$$

Sovitteet $\hat{y}_k = \mathbf{x}'_k \hat{\boldsymbol{\beta}} + \hat{u}_{0d}, \quad k \in U$

Osajoukkojen totaalien GREG-estimaattori

Totaaliparametrit $T_d = \sum_{U_d} y_k$, $d = 1, \dots, D$

Tulosmuuttuja y jatkuva tai binäärinen

GREG-estimaattorit $\hat{t}_{dGREG} = \sum_{k \in U_d} \hat{y}_k + \sum_{k \in S_d} a_k \hat{e}_k$

missä $a_k = 1/\pi_k$ (asetelmapainot)

$\hat{e}_k = y_k - \hat{y}_k$ (jäännöstermit eli residuaalit)

HUOM: Sovitteet \hat{y}_k lasketaan kulloisenkin mallin avulla

Kalibrointi *Calibration*

■ Mallista vapaa kalibrointi

Model-free calibration

- Särndal C.-E., Swensson B. and Wretman J. (1992). *Model Assisted Survey Sampling*. Springer.
- Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology* **33**, 99-119.

■ Malliavusteinen kalibrointi

Model calibration

- Lehtonen R., Särndal C.-E. and Veijanen A. (2008). Generalized regression and model-calibration estimation for domains: Accuracy comparison. Invited paper, BNU Workshop, August 2008, Kuressaare.



Mallista riippumaton kalibrointi

- Laskenta
 - SAS-makro [CLAN](#) (SCB, Ruotsi)
 - Andersson, C. and L. Nordberg (1998). A User's Guide to CLAN 97 - a SAS-program for computation of point- and standard error estimates in sample surveys, Statistics Sweden.
 - SPSS-ohjelma [g-CALIB](#) (Belgian tilastovirasto)
 - SAS-makro CALMAR (INSEE; Ranska)
 - Bascula (CBS, Hollannin tilastovirasto)
- [VLISS](#)-esimerkki
- Ks. [jaettu moniste](#)



Malliavusteinen estimointi ja kalibrointi

- Materiaali
 - [Lehtonen and Pahkinen](#) (2004) Section 3.3
 - Ratio estimation of population total
 - Regression estimation of totals
 - [Tekninen yhteenveto 2](#)
 - Jaettu paperi
 - VLISS
 - [Training Key 101](#)
 - Regression estimation
 - Monte Carlo simulation
 - [Training Key 104](#)
 - Calibration of weights



TEKNINEN YHTEENVETO 2 Malliavusteinen estimointi

Regressioestimointi

Oletetaan lineaarinen regressiomalli (superpopulaatiomalli)

$$y_k = \alpha + \beta z_k + \varepsilon_k, \quad V(y_k) = \sigma^2 \text{ (varianssi)}$$

Äärellisen perusjoukon vastineet parametreille α ja β ovat A ja B

A ja B estimoidaan painotetulla PNS-menetelmällä (WLS)

SRS-tilanne:

Kulmakertoimen B estimaattori $\hat{b} = \hat{s}_{yz} / \hat{s}_z^2$

Vakioparametrin A estimaattori $\hat{a} = \bar{y} - b\bar{z}$



Kokonaismäärän T regressioestimaattori:

$$\hat{t}_{reg} = N(\bar{y} + \hat{b}(\bar{Z} - \bar{z})) = \hat{t} + \hat{b}(T_z - \hat{t}_z), \text{ missä}$$

$$\hat{t}_{HT} = \hat{t} = N \sum_{k=1}^n y_k / \pi_k = N \sum_{k=1}^n y_k / n = N\bar{y} \text{ on tulosuuttujan } y$$

kokonaismäärän $T = \sum_{k=1}^N Y_k$ HT-estimaattori (SRSWOR),

$$\hat{t}_z = N \sum_{k=1}^n z_k / n = N\bar{z} \text{ on apumuuttujan } z \text{ kokonaismäärän}$$

$$T_z = \sum_{k=1}^N Z_k \text{ HT-estimaattori (SRSWOR)}$$

$$\bar{Z} = T_z / N$$

Tarvittava lisätieto: Apumuuttujan kokonaismäärä T_z



Asetelmavarianssi (likimääräinen)

$$V_{SRS}(\hat{t}_{reg}) \cong N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) S_y^2 (1 - \rho_{yz}^2),$$

missä $\rho_{yz} = S_{yz} / S_y S_z$ on muuttujien y ja z perusjoukon korrelaatio

Varianssiestimaattori (yksi vaihtoehto)

$$\hat{v}_{SRS}(\hat{t}_{reg}) = N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) \hat{s}_y^2 (1 - \hat{\rho}_{yz}^2)$$



Monimuuttujainen regressiomalli

$$y_k = \beta_0 + \beta_1 z_{1k} + \beta_2 z_{2k} + \dots + \beta_p z_{pk} + \varepsilon_k = \mathbf{z}'_k \boldsymbol{\beta} + \varepsilon_k$$

missä

$$\mathbf{z}_k = (1, z_{1k}, \dots, z_{pk})'$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$$

Estimaattorit ja varianssiestimaattorit

Perusmuoto

$$\hat{t}_{reg} = \hat{t}_{HT} + \hat{b}_1 (T_{z_1} - \hat{t}_{z_1}) + \hat{b}_2 (T_{z_2} - \hat{t}_{z_2}) + \dots + \hat{b}_p (T_{z_p} - \hat{t}_{z_p}) \quad (1)$$



PROC SURVEYREG-muoto

$$\hat{t}_{reg} = \hat{b}_0 N + \hat{b}_1 T_{z_1} + \hat{b}_2 T_{z_2} + \dots + \hat{b}_p T_{z_p} \quad (2)$$

```
proc surveyreg data=Sample total=32;  
model UE91=HOU85 URB85 / solution;  
weight SamplingWeight;  
estimate "UE91 Total"  
Intercept 32 HOU85 91753 URB85 7 / E;  
run;
```



GREG-muoto (Generalized regression estimator)

$$\hat{t}_{reg} = \sum_{k=1}^N \hat{y}_k + \sum_{k=1}^n w_k (y_k - \hat{y}_k) \quad (3)$$

missä $\hat{y}_k = \mathbf{z}'_k \hat{\mathbf{b}}$

HUOM: Termi $\sum_{k=1}^N \hat{y}_k$ Synteettinen (malliperusteinen)

totaalin estimaattori

Termi $\sum_{k=1}^n w_k (y_k - \hat{y}_k)$ Harhan korjaustermi

(jäännöstotaalin HT-estimaattori)



Kalibrointiestimaattori

$$\hat{t}_{reg} = \sum_{k=1}^n w_k^* y_k \quad (4)$$

missä

$w_k^* = g_k w_k$ on kalibrointipaino

$w_k = 1/\pi_k$ on asetelmapaino

g_k on g-paino alkiolle k



Varianssiestimaattorit

$$\hat{v}(\hat{t}_{reg}) = N^2(1 - n/N)(1/n)\hat{s}_{\hat{e}}^2 \quad (5)$$

missä $\hat{s}_{\hat{e}}^2 = \sum_{k=1}^n (\hat{e}_k - \bar{\hat{e}})^2 / (n-1)$

missä

$$\hat{e}_k = y_k - \hat{y}_k$$

$$\bar{\hat{e}} = \sum_{k=1}^n \hat{e}_k / n$$

$$\hat{y}_k = \mathbf{z}'_k \hat{\mathbf{b}}$$



$$\hat{v}(\hat{t}_{reg}) = N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) \left(\frac{n-1}{n-p}\right) \hat{s}_{\hat{e}^*}^2 \quad (6)$$

missä

$$\hat{s}_{\hat{e}^*}^2 = \sum_{k=1}^n (\hat{e}_k^* - \bar{\hat{e}}^*)^2 / (n-1)$$

$$\hat{e}_k^* = g_k \hat{e}_k \quad (\text{g-painotetut jäännökset})$$

$$\bar{\hat{e}}^* = \sum_{k=1}^n \hat{e}_k^* / n$$



$$\hat{v}(\hat{t}_{reg}) = N^2 (1 - n/N) (1/n) \hat{s}_y^2 (1 - \hat{R}^2) \quad (7)$$

missä

\hat{R}^2 on yhteiskorrelaatiokertoimen neliö

$$\hat{s}_y^2 = \sum_{k=1}^n (y_k - \bar{y})^2 / (n-1)$$



Kalibrointimenetelmä (Calibration)

g-painot

Tulosmuuttuja y

Kalibrointiestimaattori $\hat{t}_{reg} = \sum_{k=1}^n w_k^* y_k,$

missä

$w_k^* = g_k w_k,$ g_k on g-paino alkion k ja $w_k = 1/\pi_k$

jolle pätee $\hat{t}_{reg} = \sum_{k=1}^n w_k^* z_k = \sum_{k=1}^N Z_k = T_z$ (kalibrointiominaisuus)



Suhde-estimointi (Ratio estimation)

Jatkuva apumuuttuja z

$$g_k = \frac{T_z}{\hat{t}_z}$$

Regressioestimointi (Regression estimation)

Jatkuva apumuuttuja z

$$g_k = \frac{N}{\hat{N}} \left(1 + \frac{\bar{Z} - \bar{z}}{(n-1)\hat{s}_z} (z_k - \bar{z}) \right)$$



Jälkiositus (Poststratification)

Diskreetti apumuuttuja (luokat $g = 1, \dots, G$)

$$g_{gk} = \frac{N_g}{\hat{N}_g}, \quad g = 1, \dots, G$$

missä $\hat{N}_g = \sum_{k=1}^{n_g} w_{gk}$ on jälkiositteen g estimoitu koko



Regressioestimointi - Esimerkki

- **Example 3.13 (Lehtonen – Pahkinen 2004)**
- Yksi apumuuttuja
- Populaatio *Province'91*
- Otosaineisto
 - Aikaisemmin poimittu SRSWOR-otos, $n = 8$ kuntaa
- Strategia
 - SRS*reg



- **Taulukko**
- Province91-
perusjoukko
- N = 32 kuntaa
- Tulosmuuttuja
- UE91
- Apumuuttujat
- STR osite
- Kuntamuoto
- HOU85
- Kotitalouksien
lkm

■ Lähde: Lehtonen R. and Pahkinen E. (2004). Practical Methods for Design and Analysis of Complex Surveys. Second Edition. Wiley.

Table 2.1 The Province'91 population. Percentage unemployment (%UE) and totals of unemployed persons (UE91), labour force (LAB91), population in 1991 (POP91) and number of households (HOU85) by municipality in the province of Central Finland in 1985.

ID	LABEL	STR	CLU	%UE	UE91	LAB91	POP91	HOU85
Urban								
1	Jyväskylä	1	1	12.20	4123	33786	67200	26881
2	Jämsä	1	2	11.07	666	6016	12907	4663
3	Jämsäkoski	1	2	13.83	526	3828	6116	3019
4	Keuruu	1	2	12.84	760	5919	12707	4896
5	Saarjärvi	1	3	14.62	721	4930	10774	3730
6	Suolahti	1	5	15.12	457	3022	6159	2389
7	Äänekoski	1	3	13.17	767	5823	11595	4264
Rural								
8	Hankasalmi	2	5	15.07	391	2594	6080	2179
9	Joutsa	2	6	9.38	194	2069	4594	1823
10	Jyväskylän m.k.	2	7	11.82	1623	13727	29349	9230
11	Kannonkoski	2	4	18.64	153	821	1919	726
12	Karstula	2	4	13.53	341	2521	5594	1868
13	Kinnula	2	8	13.92	129	927	2324	675
14	Kivijärvi	2	8	15.63	128	819	1972	634
15	Konginkangas	2	3	21.04	142	675	1636	556
16	Konnevesi	2	5	12.91	201	1557	3453	1215
17	Korpilampi	2	1	11.15	239	2144	5181	1793
18	Kuhmoinen	2	2	12.91	187	1448	3357	1463
19	Eyyjärvi	2	4	11.31	94	831	1977	672
20	Laukaa	2	5	12.11	874	7218	16042	4952
21	Levernämäki	2	6	10.65	61	573	1370	545
22	Luhanka	2	6	10.34	54	522	1153	435
23	Multia	2	7	11.24	119	1059	2375	925
24	Muurame	2	1	9.79	296	3024	6830	1853
25	Peltjärvi	2	7	15.08	262	1737	3800	1352
26	Pihlupudas	2	8	13.02	331	2543	5654	1946
27	Pylkönmäki	2	4	17.98	98	545	1266	473
28	Sumainen	2	3	12.80	79	617	1426	485
29	Sivonmäki	2	1	10.28	166	1615	3628	1226
30	Toivakka	2	6	11.72	127	1084	2499	834
31	Uurainen	2	7	16.47	219	1330	3004	932
32	Vitasaari	2	8	14.16	568	4011	8641	3119
Whole province				12.65	15 098	119 325	254 584	91 753

Sources: Statistics Finland: Population Census 1985. Statistics Finland (1992): Statistical Yearbook of Finland, Volume 87. Ministry of Labour of Finland (1991): Employment Service Statistics, November 30, 1991.



Table 3.16 A simple random sample drawn without replacement from the Province'91 population prepared for regression estimation

Sample design identifiers			Auxiliary information				
STR	WGHT	Element LABEL	Study var. UE91	Variable HOU85	Model group	WGHT g-weight	Final w*
1	4	Jyväskylä	4123	26 881	1	0.2844	1.1378
1	4	Keuruu	760	4896	1	1.0085	4.0341
1	4	Saarjärvi	721	3730	1	1.0469	4.1877
1	4	Konginkangas	142	556	1	1.1057	4.6058
1	4	Kuhmoinen	187	1463	1	1.1216	4.4863
1	4	Pihlupudas	331	1946	1	1.1391	4.4227
1	4	Toivakka	127	834	1	1.1423	4.5691
1	4	Uurainen	219	932	1	1.1515	4.5562

Sampling rate = 8/32 = 0.25.



Regressioestimointi - Esimerkki

- Tehdään regressioestimointi kahdella tavalla (saadaan sama numeerinen tulos)
 - Peruskaava (1)
 - Kalibrointiestimaattori (4)
 - Varianssiestimaattori (6)
- Tulosmuuttuja y
 - UE91
- Apumuuttuja z
 - HOU85 (kotitalouksien lkm kunnassa 1985)



Regressioestimointi - Esimerkki

- Taulukko 3.16
 - Asetelmaindikaattorit vastaavat SRSWOR-tilannetta
 - Poimintasuhde on 0.25.
- Regressiomalli
 - Selitettävä muuttuja UE91
 - Selittäjä HOU85
 - Estimoitu slope (kulmakerroin) = 0.152
- Saadaan regressioestimaatti:

$$\begin{aligned}\hat{t}_{reg} &= \hat{t} + \hat{b}(T_z - \hat{t}_z) \\ &= 26440 + 0.152(91753 - 164952) = 15312\end{aligned}$$



Regressioestimointi - Esimerkki

- Sama piste-estimaatti saadaan kalibrointiestimaattorilla
 - ks. Painot Table 3.16:

$$\hat{t}_{reg} = \sum_{k=1}^8 w_k^* y_k = 15312$$

Risto Lehtonen

35



Regressioestimointi - Esimerkki

- Varianssiestimaatti
 - Kaava (6) tuottaa tulokseksi

$$\begin{aligned}\hat{v}_{srs}(\hat{t}_{reg}) &= N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) \left(\frac{n-1}{n-p}\right) \hat{s}_{e^*}^2 \\ &= 32^2 \left(1 - \frac{8}{32}\right) \left(\frac{1}{8}\right) \left(\frac{8-1}{8-2}\right) 61.24^2 = 648^2\end{aligned}$$

Risto Lehtonen

36



Regressioestimointi - Esimerkki

- Laskenta
 - SAS Procedure SURVEYREG
 - Laskenta: Kaava (2)

PROC SURVEYREG-muoto

$$\hat{t}_{reg} = \hat{b}_0 N + \hat{b}_1 T_{z_1} + \hat{b}_2 T_{z_2} + \dots + \hat{b}_p T_{z_p} \quad (2)$$

Risto Lehtonen

37



Proc SURVEYREG

```
**Regression Estimation;  
proc surveyreg data=Sample total=32;  
title2 "Regression estimation for the  
total of UE91, auxiliary variable  
HOU85";  
model UE91=HOU85 / solution;  
weight SamplingWeight;  
estimate "UE91 Total"  
Intercept 32 HOU85 91753 / E;  
run;
```

Risto Lehtonen

38



Proc SURVEYREG

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	42.6546808	22.1860968	1.92	0.0960
HOU85	0.1520142	0.0007745	196.29	<.0001



Proc SURVEYREG

Coefficients of Estimate "UE91 Total"

Effect	Row 1
Intercept	32
HOU85	91753

Analysis of Estimable Functions

Parameter	Estimate	Standard Error	t Value	Pr > t
UE91 Total	15312.7108	648.160289	23.62	<.0001



Regressioestimointi - Esimerkki

- *Kaksi apumuuttujaa*
 - Populaatio *Province'91*
 - Otosaineisto
 - Aikaisemmin poimittu SRSWOR-otos, $n = 8$ kuntaa
 - Strategia
 - SRS*reg
- Tulosuuttuja y
 - UE91
- Apumuuttujat z_1 ja z_2
 - HOU85 (kotitalouksien lkm kunnassa 1985)
 - URB85 (1 = kaupungit, 0 = muut kunnat)



Regressioestimointi - Esimerkki

- Regressiomalli
 - Tulosuuttuja y : UE91
 - Apumuuttujat
 - z_1 : HOU85
 - z_2 : URB85

$$y_k = \beta_0 + \beta_1 z_{1k} + \beta_2 z_{2k} + \varepsilon_k = \mathbf{z}'_k \boldsymbol{\beta} + \varepsilon_k$$



Regressioestimointi - Esimerkki

■ Laskenta

- Peruskaava (1)
- GREG-muoto (3) (Table 3.17)

$$\begin{aligned}\hat{t}_{reg} &= \hat{t}_{HT} + \hat{b}_1(T_{z_1} - \hat{t}_{z_1}) + \hat{b}_2(T_{z_2} - \hat{t}_{z_2}) \\ &= 26440 + 0.14956(91753 - 164952) \\ &\quad + 68.107(7 - 12) = 15152\end{aligned}$$

Risto Lehtonen

43



Regressioestimointi - Esimerkki

■ GREG-kaava (Table 3.17)

- Tuottaa saman numeerisen tuloksen (15152) kuin (1)
- Varianssiestimaatti lasketaan kaavalla (6)

GREG-muoto

$$\hat{t}_{reg} = \sum_{k=1}^{32} \hat{y}_k + \sum_{k=1}^8 w_k (y_k - \hat{y}_k) = 15152$$

missä

$$\begin{aligned}\hat{y}_k &= \mathbf{z}'_k \hat{\mathbf{b}} \\ \mathbf{z}_k &= (1, z_{1k}, z_{2k})' \\ \hat{\mathbf{b}} &= (\hat{b}_0, \hat{b}_1, \hat{b}_2)'\end{aligned}$$

Risto Lehtonen

44



Regressioestimointi - Esimerkki

■ Laskenta

■ Peruskaava (1)

- Lisäinfo: Perusjoukon totaalityöt HOU85 ja URB85

■ GREG-kaava (3)

- Lisäinfo: Yksikötason tiedot perusjoukon kaikilta alkiolta

■ SAS-laskenta



Proc SURVEYREG

```
**Regression Estimation;  
proc surveyreg data=Sample total=32;  
title2 "Regression estimation for the  
total of UE91, auxiliary variable  
HOU85 and URB85";  
model UE91=HOU85 URB85 / solution;  
weight SamplingWeight;  
estimate "UE91 Total"  
Intercept 32 HOU85 91753 URB85 7 / E;  
run;
```



Proc SURVEYREG

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	29.7768913	19.7517828	1.51	0.1754
HOU85	0.1495578	0.0023199	64.47	<.0001
URB85	68.1072704	62.7319985	1.09	0.3136

Risto Lehtonen

47



Proc SURVEYREG

Coefficients of Estimate "UE91 Total"

Effect	Row 1
Intercept	32
HOU85	91753
URB85	7

Analysis of Estimable Functions

Parameter	Estimate	Standard Error	t Value	Pr > t
UE91 Total	15151.9849	568.987386	26.63	<.0001

Risto Lehtonen

48



Table 3.18 Estimates for the population total of UE91 under different estimation strategies: an SRSWOR sample of eight elements drawn from the Province'91 population.

Estimation strategy	Estimator	Estimate	s.e	deff
Design-based				
SRSWOR		26 440	13 282	1.0000
SRSWR		26 440	15 095	1.2917
Design-based model-assisted				
Poststratified estimator	SRS*pos	18 106	6021	0.3323
Ratio estimator	SRS*rat	14 707	892	0.0045
Regression estimator	one z-variable SRS*reg,1	15 312	648	0.0020
	two z-variables SRS*reg,2	15 152	569	0.0018



Kirjallisuutta

- OPPIKIRJOJA
- Särndal C.-E., Swensson B. and Wretman J. (1992). Model Assisted Survey Sampling. New York: Springer.
- Lehtonen R. and Pahkinen E. (2004). Practical Methods for Design and Analysis of Complex Surveys. 2nd Edition. Chichester: Wiley.
- ARTIKKELEITA
- Lehtonen R., Särndal C.-E. and Veijanen A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, 29, 33–44.
- Lehtonen R., Särndal C.-E. and Veijanen A. (2005). Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition*, 7, 649–673
- Lehtonen R. and Veijanen A. (2009). Design-based methods of estimation for domains and small areas. Chapter 31 in Rao C.R. and Pfeffermann D. (Eds.). *Handbook of Statistics. Sample Surveys: Inference and Analysis. Vol. 29B*. New York: Elsevier.
- Lehtonen R. and Veijanen A. (1998). Logistic generalized regression estimators. *Survey Methodology*, 24, 51–55.